The Trunk Control Measurement Scale: reliability and discriminative validity in children and young people with neuromotor disorders

Marsico, Petra ; Mitteregger, Elena ; Balzer, Julia ; van Hedel, Hubertus J A

Abstract: AIM: This study investigated the intra- and interrater reliability of the Trunk Control Measurement Scale (TCMS) German version, with its subscores, in children with neuromotor disorders. Further, the discriminative validity of the TCMS was assessed by comparing the TCMS scores with the Functional Independence Measure for children. METHOD: Bland-Altman analyses and intraclass correlation coefficients were applied to investigate reliability. The discriminative ability of the TCMS was evaluated with receiver operating characteristics. RESULTS: Ninety children (mean age 11y 5mo; range 5y-18y 11mo) participated for the reliability, and 50 for the discriminative validity study. The reliability proved to be excellent (intrarater: bias=0.57 points, 95% confidence interval [CI] -3.71 to 4.85; interrater: bias=-0.31 points, 95% CI -5.77 to 5.10). A change in the TCMS total score of six points (10%) can be considered a true change. The TCMS subscores appeared to be clinically relevant because children with less than around 80% of the static balance score, less than 55% of the dynamic reaching score, or less than around 35% of the selective movement control score needed support for daily life activities. INTERPRETATION: The TCMS is a reliable and clinically relevant assessment for children aged 5 years and older with different neurological impairments.

DOI: https://doi.org/10.1111/dmcn.13425

Originally published at:
Marsico, Petra; Mitteregger, Elena; Balzer, Julia; van Hedel, Hubertus J A (2017). The Trunk Control Measurement Scale: reliability and discriminative validity in children and young people with neuromotor disorders. Developmental Medicine and Child Neurology, 59(7):706-712.
DOI: https://doi.org/10.1111/dmcn.13425

Trunk Control Measurement Scale: Reliability and discriminative validity in children and youths with neuro-motor disorders

Petra Marsico [*] [a,b], Elena Mitteregger [c,d], Julia Balzer [a,b], Hubertus J.A. van Hedel [a,b]

[a] *Paediatric Rehab Research Group, Rehabilitation Centre, University Children's Hospital Zurich, Muehlebergstrasse 104, CH-8910 Affoltern am Albis, Switzerland*
[b] *Children's Research Centre, University Children's Hospital Zurich, Steinwiesstrasse 75, CH-8032 Zurich, Switzerland*
[c] *Zurich University of Applied Studies, ZHAW, Institute for Physiotherapy, Technikumstrasse 71, CH-8401 Winterthur, Switzerland*
[d] *Children Therapy Centres of the Foundation Regional group Zurich (Stiftung RgZ), Rautistrasse 75, 8048 Zürich, Switzerland*

Elena Mitteregger: elena.mitteregger@zhaw.ch

Julia Balzer: Julia.balzer@kispi.uzh.ch

Hubertus van Hedel: hubertus.vanhedel@kispi.uzh.ch

*Corresponding author:

Petra Marsico, PT
Paediatric Rehab Research Group
Rehabilitation Centre for Children and Adolescents
University Children's Hospital Zurich
Mühlebergstrasse 104
CH-8910 Affoltern am Albis
Switzerland
Phone: +41 (0)44 762 52 97

E-mail address: petra.marsico@kispi.uzh.ch

Author Disclosure Statement: No competing financial interests exist.

Word count manuscript: 3128

Word count abstract: 200

ABSTRACT

AIM: This study investigated the intra-rater and inter-rater reliability of the Trunk Control Measurement Scale (TSMC) with its sub-scores, in children with neuro-motor disorders. Further, the discriminative validity of the TCMS was assessed by comparing the TCMS scores with the Functional Independence Measure for children (WeeFIM®).

METHOD: Bland- Altman analyses and Intraclass correlation coefficients were applied to investigate reliability. The discriminative ability of the TCMS was evaluated with receiver operating characteristics (ROC).

RESULTS: Ninety children (mean age 11y 5mo, range 5-19 years) participated for the reliability, and 50 for the discriminative validity study. The reliability proved to be excellent (intra-rater: bias = 0.57 points, 95%CI: -3.71 to 4.85 and inter-rater: bias = -0.31, 95%CI -5.77 to 5.10). A change in the TCMS total score of 6 points (10%) can be considered a true change. The TCMS sub-scores appeared to be clinically relevant because children with less than around 80% of the static balance score, less than 55% of the dynamic reaching score or less than around 35% of the selective movement control score needed support for daily life activities.

INTERPRETATION: The TCMS is a reliable and clinically relevant assessment for children with different neurological impairments aged five years and older.

Shortened form of title: TCMS: reliability and clinical relevance

What does this paper adds

- Confirms the reliability of the TCMS in children with various neuro-motor disorders
- Adds information on the measurement errors for the three sub-scores of the TCMS
- Adds subgroup analysis of children aged five to eight years
- The results allow interpreting longitudinal changes
- Shows the clinical relevance of trunk control for independence in daily living

The central role of trunk control on function and activity in children and youths with neuro-motor disorders is currently discussed in the literature.[1,2] Often, postural control is impaired in persons with a neurological condition and this can affect trunk stability, mobility and selectivity of the upper limbs.[2–5] In clinical practice, we address these impairments with different approaches, such as balance training during physical therapy or hippotherapy. To assess the efficacy of such interventions, we need valid, reliable, and responsive outcome measures to assess (changes in) trunk control. In a recent review, the psychometric properties of sitting balance measures for children with cerebral palsy (CP) were discussed.[6] Seven outcome assessments were included: the Pediatric Reach Test, the Level of Sitting Scale, the Sitting Assessment for children with neuro-motor Dysfunction, the Segmental Assessment of Trunk Control, the Sitting Assessment of Children with neuro-motor Dysfunction, the Trunk Impairment Scale and the Trunk Control Measurement Scale (TCMS).[6] With the TCMS, an objective outcome measure is nowadays available to score the trunk ability in sitting. The TCMS consists of static and dynamic sitting balance; the latter is divided into selective movement control and dynamic reaching.[7] This assessment tool shows good relative reliability in children with spastic CP aged eight to 15 years.[7] Information on absolute measurement errors is needed to interpret whether observed changes can be considered “true” or not. While the same study investigated also the absolute measurement error of the total score (the Smallest Detectable Difference (SDD) was 4.66 for the intra-rater reliability, and 5.47 for the inter-rater reliability), such information is missing for the sub-scores of the TCMS. Furthermore, these results were obtained in a relatively small group of children (n=26) and were restricted to children with spastic CP. In clinical practice, impaired trunk control can be observed also in younger children and those with other neurological diagnoses (e.g. dyskinetic or ataxic CP and other neuro-motor disorders, such as acquired brain injuries, myelomeningocele or different syndromes), so it is important to investigate whether these findings also apply to a broader group.

Besides the reliability of the TCMS, we were also interested to learn how much trunk control a child needs to be independent for mobility and self-care in daily life, as the child’s independence in daily living is important for the children and their families. Therefore, we also investigated the discriminative validity by comparing the TCMS and its sub-scores to the mobility and self-care domains of the Functional Independence Measure for children [WeeFIM®]. [8,9]

METHODS

Participants

Children were recruited from the in- and out-patient setting of the Rehabilitation Centre for Children and Adolescents of the University Children’s Hospital Zurich in Affoltern am Albis and the Children Therapy Centres of the Foundation Regional group Zurich (Stiftung RgZ). Inclusion criteria were: neurological diagnosis such as CP (GMFCS I-IV), acquired brain injury (ABI), spinal cord injury and the age of five to 19 years. The children needed to have the ability to follow easy-to-understand instructions. Exclusion criteria were: surgery or botulinum toxin injection within the last three months and pain or medical restriction for weight bearing. Parents and adolescents aged 15 years and above signed an informed consent form. Children below the age of 15 years agreed to participate in this study. The study was approved by the ethics committee of the Canton of Zurich, complied with the Declaration of Helsinki and followed the guidelines of good clinical practice. The aim was to collect data on at least 50 children. According to the COSMIN group, a sample size over 50 is considered as good.[10]

Measures

The TCMS has been translated and validated in the Korean and German language.[11,12] We performed the TCMS according to the German description of Mitteregger and colleagues.[12] The TCMS total score and the three sub-scores static were evaluated. The maximum value for the total TCMS is 58 (20 points for the category static sitting balance, 28 points for selective movement control and 10 for dynamic reaching). A higher TCMS scores indicates a better performance in trunk control.
The TCMS score correlates well with the Gross Motor Function Classification System (GMFCS) and the Gross Motor Function Measure (GMFM).[4,7,12] While the GMFCS and the GMFM provide information about gross motor function and capacity, respectively, they do not assess independence in daily life activities.[13,14] To assess daily life independence, we had a trained research nurse who routinely assessed the WeeFIM® values of in-patients. The WeeFIM® is an 18 item tool to observe daily life performance and independence in children. The assessment is suitable for application in all children and adolescents whose functioning level lies below that of a typically developed peer of 7 years old (i.e. a typically developed seven years old child should reach the maximum score of the WeeFIM). In additional, it can be applied to children with developmental disabilities aged six months to 21 years. The WeeFIM consists of three categories: self-care (8 items), mobility (5 items), and cognition (5 items). Each item is rated on a seven-point ordinal scale. While a score of one indicates that total assistance is required, a score of seven indicates complete independence. We were interested in the self-care and mobility categories, because we expected that trunk control should be relevant for these categories. To assess the self-care domain, three of eight items were selected: eating, dressing upper body and dressing lower body. Grooming, bathing, toileting, bladder, and bowel function were excluded, because these measures do not solely depend on voluntary motor function (e.g. modifications in the environment or autonomic bladder or bowel control). For the mobility section, all items were included: transfer to a chair or wheelchair, transfer to the toilet, transfer to the bath or shower, mobility such as walking or by means of a wheelchair, and walking stairs.

Assessment procedure

Two physiotherapists with more than ten years of experience in treating children and adolescents with neurological disorders tested all children in a quiet room. The child sat on a movable bench with the feet unsupported. Orthoses, shoes, and socks were taken off. The testing lasted 20 to 30 minutes. The testing was recorded by video in the frontal plane except for items 2, 6, 7, and 12 (these items were recorded in the sagittal plane for better analysis). For the intra-rater reliability, the TCMS tests were scored by the same rater, first directly after the original test situation from the video and the second time from the video at least two months later. For the inter-rater reliability, both therapists (rater A and B) used the video recordings to score the TCMS after more than two months after the assessment was conducted. These ratings were compared.

Data analysis

We assumed that if the TCMS could differentiate well (i.e. a good discriminative validity) between children being dependent versus independent in self-care and mobility, this could indicate the clinical relevance of the TCMS. Therefore, we dichotomized the WeeFIM® self-care (3 items) and mobility (5 items) subcategories scores. To be defined as independent the child had to reach a value equal or above 5 in each included Item, i.e. children who were able to perform the activity without personal assistance (from supervision until complete independence, i.e. WeeFIM item scores of 5 to 7) were grouped as “independent”. In contrast, children who could not perform the activity (i.e. WeeFIM item score of 0) or needed assistance to perform the activity (total assistance until minimal assistance, i.e. WeeFIM item score of 1 to 4)

were grouped as “dependent”. We included the scores made by rater A for these comparisons.

Statistical analysis

We performed statistical calculations with SPSS (IBM SPSS Statistics 19, Chicago, IL, USA). All variables were visually inspected for normal distribution (histogram) and skewness and kurtosis were analysed.

For the relative inter-rater reliability analysis of the TCMS total and sub-scores, Intraclass Correlation Coefficients (ICC) and 95% Confidence Intervals (95% CI) with the two-way random effect model were calculated.[15] The ICC was calculated as follows: ICC = $Ãs^2$/ ($Ãs^2$+ $Ãt^2$ + $Ãe^2$). Thereby, $Ãs^2$ reflects the between-subjects' effect, $Ãt^2$ the amount of trial effect or systematic error (between the two ratings) and $Ãe^2$ the amount of residual variance (i.e. random error variance).[16] For the interpretation, the following benchmarks were used: ICCs higher than 0.90: very high reliability; 0.70-0.89: high; 0.50-0.69: moderate; 0.26-0.49: low; and lower than 0.25: poor reliability.[17]

To quantify the absolute reliability, the standard error of measurement (SEM) and the smallest detectable difference (SDD) were calculated, with the following formulae; SEM = •($Ãt^2$ + $Ãe^2$) and SDD = 1.96 ×•2 × SEM.[18] We expressed the SDD also as a percentage of the maximum score. Additionally, for visual interpretation, we provided Bland-Altman plots.[19] With the Blind-Altman plots we show the bias (i.e. the mean difference between the two ratings) and the limits of agreement (LOA = bias plus or minus two times the SD). The upper LOA and the lower LOA mark the 95% CI. As most TCMS studies did not evaluate psychometric properties of the TCMS in children below 8 years, we performed separate analyses for this group.

To estimate whether TCMS total and subcategories scores could distinguish between children who were independent in mobility or self-care, receiver operating characteristics (ROC) were performed. For each measure, corresponding cut-off levels were calculated, based on the Youden-Index (this is the highest sensitivity + specificity -1).[20] The area under the curve (AUC) was taken as an indicator of the accuracy and interpreted by the following benchmarks: AUC higher equal 0.90: outstanding; 0.80-0.89: excellent and 0.70-0.79: acceptable In general, alpha was set at 0.05. [21]

RESULTS

Participants

Ninety children with mean age 11years 5 months (range 4y 0mo to 18y 11mo; 44 girls, 46 boys) were included in the reliability study. Twenty-one children were below eight years, with a mean age of 6 years 3 months (SD 0.7y 4mo). The distribution of diagnoses and GMFCS levels of both groups (reliability and clinical importance) are shown in Table 1. WeeFIM data were available from the 50 in-patients.

Reliability

In two children, item 15 (reach across the midline – sub-score dynamic reaching) was not video recorded. For these children, the reliability analyses of the dynamic reaching and total TCMS score were based on items 13 and 14. All other datasets were complete. Rater A performed the assessment in 54 children, and rater B in 36 children.

The relative intra-rater and inter-rater reliability were very high, with ICC>0.90 for the total, and all sub-scores (Table 2). The absolute measurement errors of the total and sub-scores were calculated. The SDD of the total score of the TCMS remained below 10% of the maximal possible score (Table 2). The Bland-Altman plots represented the bias and the limits of agreement (95%CI) both for the intra-rater (Fig. 1a) as well as

inter-rater (Fig. 1b) reliability. The 95%CIs were well in agreement with the SDD values.
For the intra-rater reliability, the children younger than eight years had a bias of 0.67 (95%CI -4.37 to 6.09) while the children aged eight years and older showed a bias of 0.55 (95%CI -3.53 to 4.63). For the inter-rater reliability, the children younger than eight years had a bias of -0.83 (95%CI -6.54 to 4.88) while the children aged eight years and older had a bias of -0.16, (95%CI -5.48 to 5.16). A separate Bland-Altman plot of the children with ataxic and dyskinetic CP, ABI, and other neurological disorders (n=41) is included in the Appendix. All results were in line with those from the whole group.

Discriminative validity
The mean age of the 50 children were the WeeFIM was collected was 11years 5 months with a range of 5 years 0 month to 18 years 11 month. These children showed a mean TCMS total score of 32 (SD 16), with a minimum score of 1 and a maximum score of 57. Results of the subcategories were the following: static sitting balance - mean of 14 (SD 5), range 1 to 20; selective movement control - mean of 12 (SD 8), range 0 to 27; dynamic reaching - mean of 6 (SD 3), range 0 to 10 points. The mean WeeFIM® score amounted to 95 (SD 35) and ranged between 27 and the maximum of 126. All ROCs showed a significant AUC (Fig. 2). The TCMS total score could discriminate with best combined sensitivity and specificity between dependent and independent children at a cut-off value of 30.5 (i.e. 52.6% from the maximal possible TCMS score), both for self-care and mobility WeeFIM sub-scores (see also Fig.2). The relative cut-off values that could discriminate best between children with dependent versus independent self-care amounted to 77.5%, 55.0% and 30.4% for static sitting, dynamic reaching and selective movement control, respectively. Similar values were obtained to discriminate between children with dependent versus independent mobility: 82.5%, 55.0% and 41.1%. Eight children were, according to the WeeFIM® self-care domain, dependent, despite the fact that their total TCMS score exceeded the cut-off level of 30.5 points. In the mobility part, there were two children with ABI, one with myelomeningocele and two with spastic CP, and one with dyskinetic CP (three children were also rated as dependent in self-care even though trunk control was good), who were rated as dependent, and the fact that their TCMS total score exceeded the cut-off value of 30.5 points. On the contrary, three children (self-care) and two children (mobility) were rated as independent, despite that their TCMS total score was below the cut-off value. These were children diagnosed with spastic CP (two children) and one child with MMC out of the group younger than eight years (in self-care and mobility).

DISCUSSION

The aim of this study was to determine the relative and absolute reliability and the discriminative validity of the German TCMS version in children with neuro-motor disorders aged five to 19 years. The ICCs indicate a very high intra- and inter-rater reliability. The results were in line with those of previous studies [7,11] More importantly, the small SEM and SDD values indicate a good absolute agreement. As we also presented the SEM and SDD values of the three sub-scores, we can better interpret the results of a recently published study.[22] In their study, they investigated the change scores of the TCMS after three weeks of intensive physical training. All their changes lied under the SDD of the sub-scores, and therefore they could be a result of measurement error and not due to a real improvement in trunk control. As the measurement errors were smaller when tested by the same rater, we recommend that if the TCMS is used to picture clinical progress, the same rater should assess the child. The TCMS was reliable in children with highly impaired trunk control as well as in children with slight impairments. Even in children younger than eight years, the test

showed reliable results (see Fig. 1; distribution of mean scores). The TCMS could be applied reliable to children with neurological diagnosis other than spastic CP in our study (Appendix). The children with dyskinetic CP had more difficulty to reach higher points in the selective trunk control part, as they show many involuntary movements and the selectivity is markedly impaired. As we scored the TCMS using video recordings, we did not account for the natural variability regarding performance, motivation, mood, and compliance, as this would be conducted in a test-retest design. While these factors could reduce the reliability of the TCMS, especially in younger children, our results are still in line with those from Heyrman et al., who used a repeated measurement procedure with a mean time of 10 days in between the assessments.[7] According to our results, the TCMS score should increase with at least 6 points or 10% of its total score to account for the measurement error and be considered a true change. In future studies the responsiveness and minimal clinically important difference should be determined. This additional information would help the therapists to better interpret the test results of the TCMS.

The TCMS could differentiate well between children who are independent in self-care and mobility (measured with the WeeFIM®) and those who are not, indicating a good discriminative validity. The excellent AUC's were significant meaning that the discrimination of the TCMS was above chance. This was the case for the TCMS total score and for all three sub-scores. Interestingly, the relative cut-off values of the static sitting balance, dynamic reaching, and selective movement control sub-scores decreased consecutively. On the one hand, this finding could be influenced by the scaling of the sub-scores, which appears the roughest for the static sitting balance sub-score and the finest for the selective movement control sub-score. On the other hand, the highest ceiling effect occurs in the static sitting balance sub-score, while no child achieved the maximal selective movement control sub-score. What we experienced practically was that the items of the selective movement control were more difficult to perform than the items of the other parts. As the WeeFIM® does not consider quality of movement, we assume that even if children would have a moderately impaired selective movement control of the trunk, these children could still be independent in self-care and mobility, as measured with the WeeFIM®. We suggest that clinicians should be careful in using these cut-off values as they rely on various methodological decisions, such as the dichotomization of the WeeFIM in 'dependent' and 'independent'. Changing the definitions or using another measure for dichotomization would likely result in different TCMS cut-off values and this should be considered when interpretation our results.

Typically developing children can reach functional independence with the WeeFIM items we included in this study at the age of about five years, however, the majority of the children below eight years in our study were rated as dependent. This could be due to the sensorimotor impairments which might have contributed to a reduction or delay in self-dependence in these children.

This study has some methodological considerations. For the reliability analysis, we reanalysed a previously assessed TCMS using the video recordings. Our reliability analyses, therefore, do not include variability caused by a second performance of the child or adolescent. Concerning the discriminative validity analysis, this study does not investigate causality between trunk control and self-dependence. The cut-off values showed a high sensitivity and specificity, which does indicate that certain TCMS subcategories can differentiate well between children with different levels of self-dependence, but we did not integrate other factors that could contribute to daily activities such as cognition, global motor function, or muscle strength.

CONCLUSION

The reliability of the TCMs was high in this group of children aged five to 19 years with neuro-motor impairments. When using the TCMS total score as an outcome measure, the change should exceed 6 points to be higher than the measurement error. With cut-off values of the TCMS we can discriminate between the children who are independent or dependent in self-care and mobility. The relative cut-off values were different between the TCMS sub-scores which might reflect the differences in scaling of these sub-scores.

Conflict of interest statement

The authors alone are responsible for the content and writing of the paper. The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The authors wish to thank the children and their parents for their participation in this study. Highly appreciated is also the support of the physiotherapists and their help for the organization of the measurements. This work was in part sponsored by a grant from the Mäxi-Foundation. Julia Balzer was paid from grants from the PhysioSwiss, Switzerland; Physiotherapy Science Foundation, Switzerland; and the Swiss National Science Foundation (Project 32003B_156646), Switzerland.

REFERENCES

1. Curtis DJ, Butler P, Saavedra S, et al. The central role of trunk control in the gross motor function of children with cerebral palsy: a retrospective cross-sectional study. *Dev Med Child Neurol*. 2014; **57**: 351-357.

2. Saether R, Helbostad JL, Adde L, Braendvik S, Lydersen S, Vik T. The relationship between trunk control in sitting and during gait in children and adolescents with cerebral palsy. *Dev Med Child Neurol*. 2015; **57**: 344-350.

3. Brogren Carlberg E, Löwing K. Does goal setting in activity-focused interventions for children with cerebral palsy influence treatment outcome? *Dev Med Child Neurol*. 2013; **55**: 47-54.

4. Heyrman L, Desloovere K, Molenaers G, et al. Clinical characteristics of impaired trunk control in children with spastic cerebral palsy. *Res Dev Disabil*. 2013; **34**: 327-334.

5. Heide JC Van Der, Hadders-algra M. Postural Muscle Dyscoordination in Children with Cerebral Palsy. 2005; **12**: 197-204.

6. Bañas BB, Gorgon EJR. Clinimetric Properties of Sitting Balance Measures for Children with Cerebral Palsy: A Systematic Review. *Phys Occup Ther Pediatr*. 2014; **34**: 313-334.

7. Heyrman L, Molenaers G, Desloovere K, et al. A clinical tool to measure trunk control in children with cerebral palsy: the Trunk Control Measurement Scale. *Res Dev Disabil*. 2011; **32**: 2624-2635.

8. Msall ME, Di Gaudio K, Rogers BT, et al. The Functional Independence Measure for Children (WeeFIM). Conceptual basis and pilot use in children with developmental disabilities. *Clin Pediatr*. 1994; **33**: 421-430.

9. Ziviani J, Ottenbacher KJ, Shephard K, Foremen S, Astbury W, Ireland P. Concurrent Validity of the Functional Independence Measure for Children (WeeFIMTM) and the Pediatric Evaluation of Disabilities Inventory in Children with Development Disabilities and Acquired Brain Injuries. *Phys Occup Ther Pediatr*. 2001; **21**: 91-101.

10. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG. Rating the methodological quality in systematic reviews of studies on measurement properties/: a scoring system for the COSMIN checklist. *Aual Life Res*. 2012; **21**: 651-657.

11. Jeon JY, Shin WS. Reliability and validity of the Korean version of the Trunk Control Measurement Scale (TCMS-K) for children with cerebral palsy. *Res Dev Disabil*. 2014; **35**: 581-590.

12. Mitteregger E, Marsico P, Balzer J, van Hedel HJA. Translation and construct validity of the Trunk Control Measurement Scale in children with brain lesions. *Res Dev Disabil*. 2015; **45-46**: 343-352.

13. Russell DJ, Avery LM, Rosenbaum PL, Raina PS, Walter SD, Palisano RJ. Improved scaling of the gross motor function measure for children with cerebral palsy: evidence of reliability and validity. *Phys Ther*. 2000; **80**: 873-885.

14. Palisano R, Rosenbaum P, Walter S, Russel DEW, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol*. 1997; **39**: 214-233.

15. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psyschological Bull*. 1979; **86**: 420-428.

16. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med*. 2002; **21**: 3431-3446.

17. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care*. 1992; **30**: 917-925.

18. De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006; **59**: 1033-1039.

19. Bland JM, Altman DG. Comparing methods of measurement/: why plotting difference against standard method is misleading. *Lancet*. 1995; **346**: 1085-1087.

20. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950; **3**: 32-35.

21. Hosmer D, Lemeshow S, Sturdivant R. *Applied Logistic Regression (eBook)*. 3. Ed. New Jersey: Wiley; 2013.

22. Pham HP, Eidem A, Hansen G, et al. Physical & Occupational Therapy In Pediatrics Validity and Responsiveness of the Trunk Impairment Scale and Trunk Control Measurement Scale in Young Individuals with Cerebral Palsy Validity and Responsiveness of the Trunk Impairment Scale and Trunk Contro. *Phys Occup Ther Pediatr*. 2016; **36**: 440-452.

Table 1

Number of participants apportioned by diagnosis and disability level according to the Gross Motor Function Classification System (GMFCS)

Analysis	GMFCS	Spastic cerebral palsy (n)	Ataxic and dyskinetic cerebral palsy (n)	Acquired brain injuries (n)	Others (n)	Total (n)
Reliability	Level I	20	6	13	2	41
	Level II	13	6	3	1	23
	Level III	12	3	0	2	17
	Level IV	4	3	1	1	9
	Total	49	18	17	6	**90**
Clinical relevance	Level I	7	1	7	1	16
	Level II	6	2	3	2	13
	Level III	3	2	2	1	12
	Level IV	7	2	3	1	9
	Total	23	7	15	5	**50**

Abbreviations: n – number of observations

Comment: The group 'others' includes four children with Myelomeningocele/Hydrocephalus, one child with Guillain-Barré-Syndrom, and one child with Hereditary Paraplegia (this child was out-patient, and is only Included in the reliability analysis).

Table 2

Intraclass correlation coefficients for the reliability with the 95% Confidence interval, Standard error of measurement and the Smallest Detectable Difference

Reliability	TCMS	ICC (95% CI)	SEM	SDD	SDD/total
Intra-rater	Static sitting balance	0.96 (0.94-0.97)***	0.61	1.70	9%
	Selective movement control	0.96 (0.94-0.98)***	1.43	3.97	14%
	Dynamic reaching	0.95 (0.93-0.97)***	0.65	1.81	18%
	Total Score	0.98 (0.97-0.99)***	1.58	4.39	8%
Inter-rater	Static sitting balance	0.96 (0.94-0.97)***	1.06	2.94	15%
	Selective movement control	0.94 (0.91-0.96)***	1.73	4.78	17%
	Dynamic reaching	0.96 (0.94-0.97)***	0.66	1.84	18%
	Total Score	0.98 (0.97-0.99)***	1.90	5.27	9%

Abbreviations: TCMS – Trunk control measurement scale; ICC – Intraclass correlation coefficient; SEM – standard error of measurement; SDD – smallest detectable difference; *** p-value <0.001

Figures legends

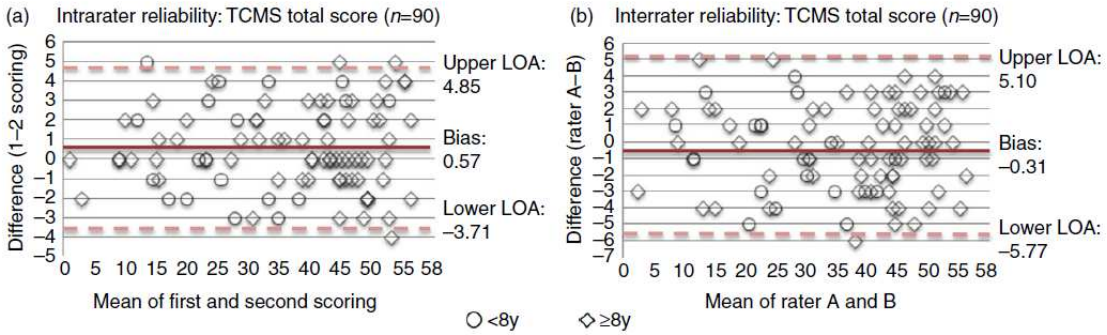


Figure 1. Intra-rater and inter-rater reliability of the TCMS total score

Intra-rater reliability A. and Inter-rater reliability B. expressed by Bland-Altman, with the bias and the limits of agreement. The diamonds represent the children of eight years and older while the circles represent the children below eight years.
Abbreviation: LOA – limit of agreement

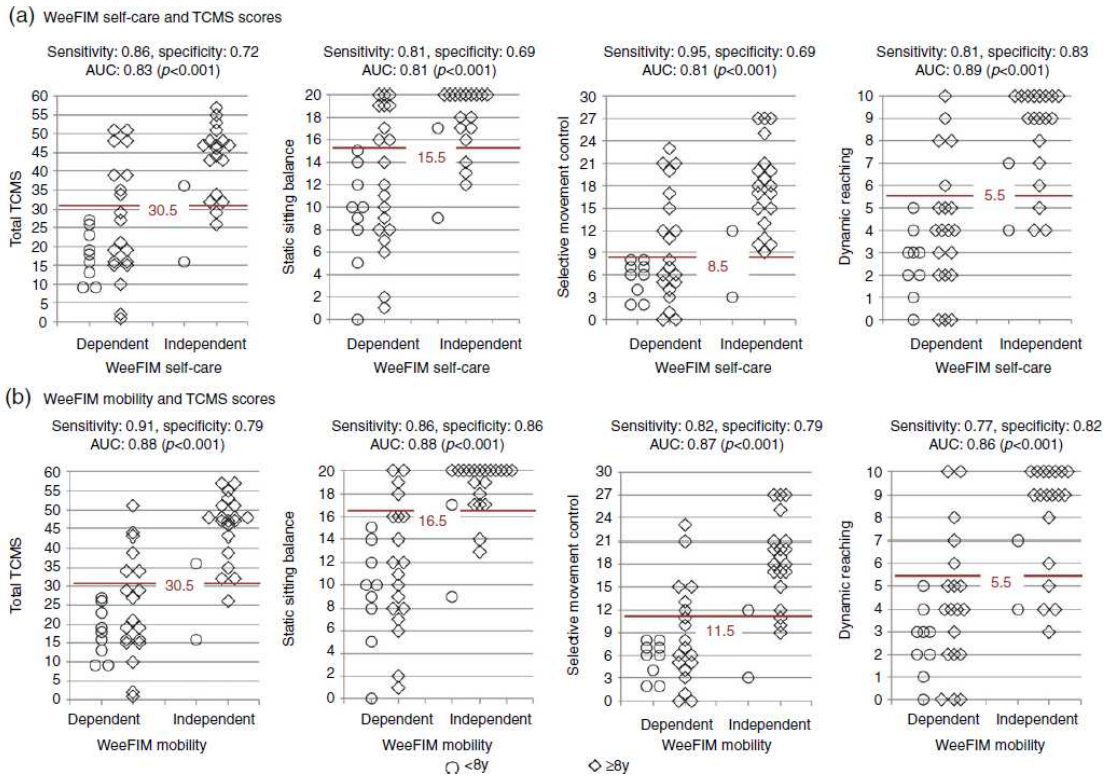
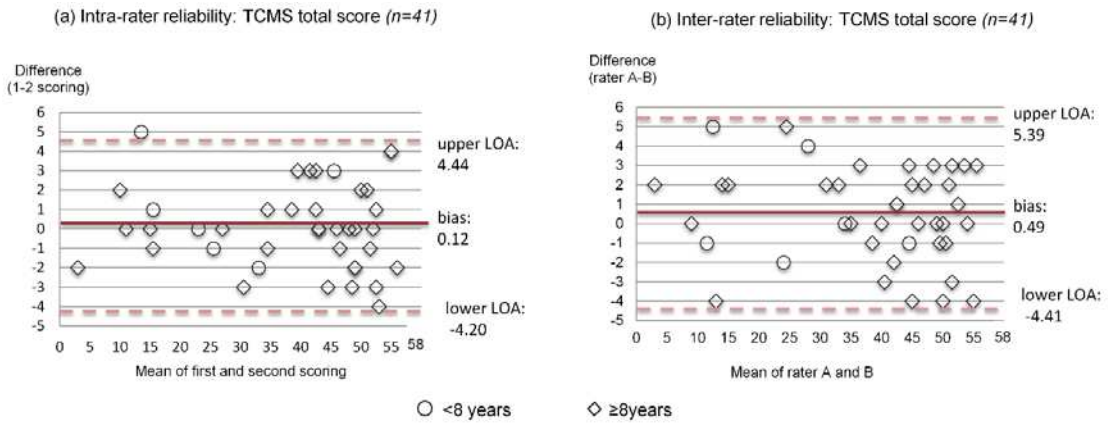


Figure 2. Discriminative validity

Presented are cut-off values, sensitivity, specificity, and area under the curve values of the TCMS scores (total score, static sitting balance, selective movement control and dynamic reaching scores) to distinguish between children who are according to the WeeFIM dependent versus independent for A. self-care domain and B. mobility. The diamonds represent the children of eight years and older while the circles represent the children below eight years.
Abbreviation: TCMS – Trunk Control Measurement Scale; WeeFIM® – Functional Independence Measure for children, AUC – area under the curve



Appendix. Intra-rater and inter-rater reliability of the TCMS total score

Intra-rater reliability of the children with ataxic and dyskinetic CP, ABI, and other neurological disorders (n=41) A. and Inter-rater reliability B. expressed by Bland-Altman, with the bias and the limits of agreement. The diamonds represent the children of eight years and older while the circles represent the
children below eight years.
Abbreviation: LOA – limit of agreement